Centre for Education Policy & Equalising Opportunities

CEPEO Working Paper No. 24-06

Occupational hazard: Inequalities in labour market mismatch

Lindsey Macmillan
UCL

Richard Murphy
University of Texas

Gill Wyness
UCL

In this paper we depart from traditional skills-based measures of occupational mismatch. Whereas skill-based measures are typically non-hierarchical, and involve comparing an individual's skills to those required by their occupation, we devise a new hierarchical method. Specifically, we create two continuous, measures of occupational quality: an 'input' measure derived from the initial qualifications of others in an occupation, and an 'output' measure derived from the realized wages of others, alongside a corresponding measure of individual ability. We use these detailed, comparable measures to examine the extent to which individuals mismatch into occupations, for the first time in the literature. We explore the nature of mismatch throughout the ability distribution, focusing on systematic differences by socio-economic status (SES) and gender. We find low SES individuals are employed in lower wage and lower qualification occupations compared to their similarly qualified peers. Meanwhile, while females match to occupation groups with higher achieving employees than males, they are employed in lower wage occupations. Educational routes between compulsory education and occupations at age 25 can explain around 33% of these SES gaps among high achievers, but persistent sizeable difference remain, conditional on all post-16 activity. By contrast the gender gap in mismatch remains stable, suggesting that education choices are not driving the differences. Instead, industry worked in accounts for most of the gender gap, though only among low achievers.

VERSION: October 2024

Suggested citation: Macmillan, L., Murphy, R., & Wyness, G. (2024). *Occupational hazard: Inequalities in labour market mismatch* (CEPEO Working Paper No. 24-06). UCL Centre for Education Policy and Equalising Opportunities. https://EconPapers.repec.org/RePEc:ucl:cepeow:24-06.

Disclaimer

Highlights

• We ask which types of individuals are more likely to “undermatch” in their occupation, by entering a lower quality occupation than might be expected given their qualifications? We use data from Next Steps to investigate this across the entire spectrum of achievement and jobs.

• We measure occupation quality according to i) the education levels of workers in that occupation (where the highest quality occupations are those with the most highly educated workers), and ii) the average earnings of the occupation (where the highest quality occupations are those with the highest earning workers)

• We find that individuals from low SES backgrounds are more likely to “undermatch”, working in occupations that are lower ranked in terms of both earnings, and education levels than those from high SES backgrounds with the same qualifications.

• In contrast, while women are more likely to work in lower paying occupations than men across the achievement distribution, they also work in occupations that have, on average, more qualified workers.

• We find that these gaps cannot be explained by prior attainment, experience, non-cognitive factors such as academic self-confidence, or a range of measures of occupational preferences. Only industry matters, for the gender gap in occupations.

Why does this matter?

Understanding which types of young people enter lower quality occupations than they could have given their qualifications has implications for both social mobility and the gender pay gap. We should target low income and female students with better careers information, such as on the occupations that match their attainment profile, and the earnings associated with different occupations.

Occupational hazard: Inequalities in labour market mismatch

Lindsey Macmillan[*,†], Richard Murphy[†,**], Gill Wyness[†,*,††]

October 2024

*UCL Centre for Education Policy and Equalising Opportunities

**University of Texas at Austin, NBER, CESifo

† IZA Institute of Labor Economics

†† Centre for Economic Performance, LSE

<u>Abstract</u>

In this paper we depart from traditional skills-based measures of occupational mismatch. Whereas skill-based measures are typically non-hierarchical, and involve comparing an individual's skills to those required by their occupation, we devise a new hierarchical method. Specifically, we create two continuous, measures of occupational quality: an 'input' measure derived from the initial qualifications of others in an occupation, and an 'output' measure derived from the realized wages of others, alongside a corresponding measure of individual ability. We use these detailed, comparable measures to examine the extent to which individuals mismatch into occupations, for the first time in the literature. We explore the nature of mismatch throughout the ability distribution, focusing on systematic differences by socio-economic status (SES) and gender. We find low SES individuals are employed in lower wage and lower qualification occupations compared to their similarly qualified peers. Meanwhile, while females match to occupation groups with higher achieving employees than males, they are employed in lower wage occupations. Educational routes between compulsory education and occupations at age 25 can explain around 33% of these SES gaps among high achievers, but persistent sizeable difference remain, conditional on all post-16 activity. By contrast the gender gap in mismatch remains stable, suggesting that education choices are not driving the differences. Instead, industry worked in accounts for most of the gender gap, though only among low achievers.

JEL classification: I20, I24, J16, J24

Keywords: mismatch, inequalities, occupational choices, education, gender

Acknowledgements

Macmillan and Wyness acknowledge funding from Nuffield Foundation grant "Inequalities in Access to Elite Occupations"

1. Introduction

The extensive literature on occupational mismatch is based on two distinct measures of mismatch. A skills-based measure, which has been used to explore how occupation skills relate to wages (Ingram and Neumann 2006), and a binary education-based measure focusing on graduates in non-graduate jobs (Dickson et al, 2022; Green and Henseke, 2016). These metrics have been used as the explanations for why some workers are more productive in an occupation than another (Guvenen et. al. 2020), or as an outcome for questions relating to social mobility (Macmillan et al., 2015, Crawford et al., 2014).

Early career choices can have a lasting influence on young people's later labour market outcomes, with graduates who enter non-graduate jobs being shown to face wage penalties and underemployment later in life (Dickson et al., 2022, Green and Henseke, 2016). Yet most of the attention to date on mismatch in the labour market has been around the coarse binary measure of over-education of graduates, or the matching of non-ordered skill groups. Due to the limitations of these metrics little is known about how young people match to occupations across the entire distribution of achievement, the magnitude of the mismatching, and importantly how this type of mismatch varies by key demographic characteristics.

In this paper, we propose a transparent empirical unidimensional measure of match that is based on the initial achievement of the worker and their peers in their occupational group. To illustrate, we operationalise it with two measures of occupational quality: an 'input' based measure derived from the initial qualifications of others in the same occupation, and an 'output' based measure derived from the realized wages of others in the occupation. This approach arises naturally from the complementarities between highly qualified individuals employed in high qualification occupations (Guvenen et al., 2020). Using this metric we are able to provide insights into the extent to which young people mismatch into occupations, and how this varies by gender and socio-economic status (SES), for the first time in the literature.

Our approach involves comparing an individual's achievement to their occupation quality. Individual's achievement is measured using their performance in exams at the end of compulsory schooling at age 16. As all individuals are required to take these examinations, this allows us to rank all young people in the national distribution, providing us with a detailed measure of initial achievement, including those who do not continue in post-compulsory schooling. Occupational quality is measured at 4-digit SOC code level, and we rank these occupational groups in two ways. First, we rank based on the average highest educational qualification of employees. Second, we rank occupations based on the average earnings of employees. This allows us to look more explicitly at the earnings implications of occupational mismatch, and its role in intergenerational income persistence. Our measure of match is simply the difference between an individual's achievement percentile rank and an occupation's percentile rank. This provides us with a detailed continuous measure of match for each individual.

We use these match measures to document SES and gender differences in occupational mismatch across the distribution of young persons' achievement. First, we present a flexible approach, plotting young person achievement deciles against occupation quality deciles. This allows direct comparisons of chosen occupational groups between individuals with the same initial qualifications. We explore inequalities in match by SES and gender across the distribution for education-based and earnings-based occupation mismatch. Second, we estimate SES and gender gaps in mismatch, and provide a mediation analysis through the introduction of demographic covariates. Third, we further explore potential mechanisms to inequalities in occupational mismatch, using detailed survey data to highlight the role of market failures and preferences in creating these inequalities.

We find persistent systematic inequalities in the match between young peoples' achievement and their occupation ranking. Students from low SES backgrounds undermatch into lower qualified and lower paid occupations across the entire distribution of achievement, relative to their similarly achieving high SES peers. At the top of the distribution, low SES young people work in occupations that are 11 percentiles lower ranked that high SES young people. We also find large gender gaps in earnings match, with low achieving women in particular working in occupations ranked 16 percentiles lower than those of low achieving men. For high achieving women, the gender gap is smaller at 3 percentiles.

While educational routes between compulsory education and occupations at age 25 can explain around 33% of these SES gradients among high achievers, a sizeable difference in undermatch remains for high achieving low SES students (8 percentiles), when taking into account all post-16 activity. The gender gap in mismatch remains stable, suggesting that education choices are not responsible for the large differences observed between men and women. Instead, the type of industry worked in can account for almost 76% of the gender gap among low achievers, although there still remains a significant difference between men and women, with low achieving women undermatching into occupations ranked 4 percentiles lower than men, even after accounting for industry. At the top of the achievement distribution, industry makes little difference to the gender gap with women consistently undermatching by around 3 percentiles. Other potential drivers of mismatch including parental occupation, motivation for occupation choice, academic self-concept, and information, advice and guidance measures all explain very little of the SES and gender gaps in occupational mismatch.

Our analysis makes several contributions to existing literatures on overeducation and education mismatch. First, we provide a simple unidimensional measure of occupational match. In contrast to much of the literature on occupational mismatch, which has focused on discrete measures of qualifications and overeducation (Flisi et al., 2017, Elias and Purcell, 2004, Steffy, 2017), ours is a continuous measure of occupational mismatch. This means we can examine mismatch across the entire distribution of achievement, including among non-graduates. It also allows us to examine the extent of mismatch. This also provides a new perspective on the studies of skill occupational mismatch, which are unable to provide rankings of occupations, rather just the skills match between an individual and an occupation. Our work builds on recent contributions to the literature on mismatch in higher education by extending this into the labour market (Campbell et al., 2022).

Second, the vast majority of the previous literature on overeducation has focused on the later outcomes of those who experience occupational mismatch, but there has been very little focus on inequalities in who mismatches with regards to occupations (Green and Henseke, 2016, Dickson et al., 2022, Mavromaras et al., 2013, McGuiness, 2006, Quinn and Rubb, 2006). Our results show that individuals from low SES backgrounds are working in less academically prestigious jobs with lower pay. This has important implications for social mobility, as it establishes that reducing educational achievement inequalities would not be sufficient to equalise wage outcomes. Similarly, that we find women tend to be in occupations where they are overqualified, and underpaid, compared to their similarly achieving male counterparts, has direct consequences for how we should tackle the gender pay gap.

The next section describes our data and empirical approach in detail. Section 3 gives an overview of our main results, including robustness tests, while Section 4 describes our analysis of potential mechanisms to disentangle the role of market failures and preferences. We end in Section 5 with a discussion and some conclusions.

2. Data and Methods

To investigate inequalities between individual quality and occupation quality we use the Next Steps cohort study, a representative sample of young people born in 1989/90 in England. The cohort were originally surveyed in 2004 at 14, and followed annually until 2010. A follow-up survey was then commissioned in 2015 to capture age 25 early labour market experiences. The survey is linked to the National Pupil Database (NPD) administrative data records giving full information on participants test scores and examination results at age 11, 16, and 18. At age 25 the survey participants' occupation is recorded using 4-digit SOC codes.

To measure individual quality, we use the total point score the individual achieved in their General Certificated of Secondary Education (GCSEs) exams at the end of compulsory secondary education. These examinations cover material from the final 2 years of schooling and students typically study up to 10 subject areas including maths, science, English, a modern foreign language, a humanities subject, and some wider curriculum options. The NPD captures a grade for each exam taken, which is allocated a points score and the sum of these is used to rank individuals. Given that we use survey data with the common issue of attrition (see Table 1), we assign individual percentiles in the survey based on their point score's position in the national distribution of point scores for every pupil in England from administrative data. Table 1 illustrates that the attrition from the survey at age 25 is as expected with those from lower SES backgrounds, men, those from ethnic minority backgrounds and lower attainers more likely to leave the survey between age 14 (wave 1) and the age 25 follow up (wave 8).

To measure occupational quality, we use an objective realised match approach[1] based on survey respondents' reported occupation at age 25 which is assigned a 4-digit Standard Occupational Classification (SOC) code. Our focus is on entry-level occupations, as is common in the literature, to minimise the impact of later labour market experiences in occupational match. Our data has the benefit of capturing a specific cohort of young people, who have recently entered the labour market, removing issues of cohort effects. There are 369 4-digit SOC codes in our sample, allowing for a very fine-graded measure of occupation. We rank occupational quality in two ways. Using the Labour Force Survey, a nationally representative survey of the employment circumstances of the UK population, we rank occupations (using the

[1] In contrast to subjective approaches (Battu et al., 2000, Green and Zhu, 2010, Baert et al., 2013), this removes the issue of measurement error based on differential reporting of match.

4-digit SOC code) based on 1) the average highest educational achievement[2], and 2) the average hourly wage[3], of all employed or self-employed workers aged 25-60[4] in the given occupations from 2014-2016. While measures of mean education have been used previously in the literature (Bauer, 2002, Elias and Purcell, 2004, for example), they have been typically used at the 1-digit or 2-digit occupation level, limiting the variation across jobs. By measuring occupational quality at the 4-digit SOC code level, and using a new metric of occupational quality, based on the labour market value of the given occupation (hourly wages), we are able to look at more detailed patterns in occupational mismatch across multiple dimensions. Figures 1 and 2 show the distribution of average education levels and average hourly wage across occupational classifications. While previous literature has divided occupational classifications into graduate and non-graduate jobs (often based on high-level SOC codes, equivalent to the left third of Figure 1 and 2), we can see that variation exists in terms of both education, and to a greater degree, earnings within broad occupation groupings.

We assign our survey respondents these rankings of occupational quality using their occupations' position in *the national ranking* for each metric, based on 335 unique occupations in our sample. There are 34 occupations that no survey members are working in at the time of the survey. The distributions of occupations vary slightly from the LFS to our sample, with a correlation of 0.8 between the proportions working in each occupation in the LFS relative to Next Steps. Appendix Figure A1 shows the relative differences in size of occupations by occupation rankings, which on the whole look broadly balanced across the distribution of occupation. While the LFS has a greater proportion of cleaners (1.8% vs 0.5% in Next Steps) and nurses (2.9% vs 1.9% in Next Steps), the Next Steps cohort have a greater proportion of sales and retail assistants (4.3% vs 2.7% in LFS), marketing associate professionals (1.5% vs 0.4% in LFS) and customer service occupations (2.0% vs 0.9% in LFS). Table 2 illustrates that we only observe occupational status for those who are in employment or self-employment at age 25, meaning that there is slightly more selection into the sample for men relative to women.

[2] For education, there are 1175 observations per occupation level on average. Our results are robust to restricting our sample to occupations with over 20 observations. Results available on request.
[3] For earnings, there are 285 observations per occupation level on average. Our results are robust to restricting our sample to occupations with over 20 observations. Results available on request.
[4] Our results are robust to restricting the rankings to survey respondents age 25-45, with spearman rank correlations of 0.99 for education-based and earnings-based occupation rankings based on 25-60 year olds and 25-45 year olds. Results available on request.

We calculate the match between survey respondents' own ranking, and both of their occupation rankings to create two measures of occupational mismatch:

1) Education match: occupation ranking (based on average education level of workers in the occupation from national distribution in the LFS) – individual ranking (based on total GCSE point score position in the national distribution in the NPD).
2) Earnings match: occupation ranking (based on average hourly pay of workers in the occupation from national distribution in the LFS) – individual ranking (based on total GCSE point score position in the national distribution in the NPD).

This gives us two continuous measures of match for each survey member, meaning that we can explore inequalities in mismatch across the entire distribution of individual achievement, instead of relying on binary graduate / non-graduate cut-offs.

Our education measure of occupational match illustrates the extent to which young people are working in occupations with similarly qualified individuals, given their achievement levels. The earnings measure of occupational match measures whether young people are working in occupations with the earnings level that we might expect, given their achievement levels. Both measures are capturing a different aspect of occupational quality. Appendix Figure A2 shows the relationship between occupation-earnings rankings and occupation-education rankings for our sample. Some occupations are highly skilled and require workers with high levels of educational qualifications, but are not well rewarded in terms of pay (for example clergy, teachers). Alternatively, some occupations are highly rewarded in the labour market, but do not require high levels of educational qualifications (for example sports players, plasterers, train drivers).

To examine inequalities in occupational match, we consider measures of the socio-economic status of young people in childhood, and differences by gender. Socio-economic status is measured in multiple different ways in the Next Steps survey, and we check the consistency of our findings across different measures. Our main measure of SES is based on the National Statistics Socio-Economic Classification (NS-SEC) of the household reference person in wave 1 of the survey when the young person is age 14. This measure combines information on the conditions of employment and employment relations of the household reference person with their occupation. This measure aims to differentiate positions in labour markets in terms of employment relations, which equate to sources of income, economic security, and advancement prospects. It also differentiates levels of authority, control and autonomy in the work place. We operationalise this measure combining the top two NS-SEC groups, the three middle groups, and the bottom two groups, into high, medium, and low parental SES. We also test the robustness of our SES gradients in occupational mismatch using a measure of parental education, which groups the highest parental qualification into high (degree or above), medium (A levels or equivalent), and low (GCSEs or below). Gender is measured based on self-reported gender of survey respondents in wave 1.

Our sample is defined as survey respondents with both a GCSE total points score from the matched NPD data and a 4-digit occupational SOC code at age 25. As discussed, this only includes respondents who are employed or self-employed at age 25. Table 1 illustrates that in the full sample there is around one third of families in each group, while in our final sample, relying on age 25 occupation reporting, 30% are from low SES families, and 37% are from high SES families. There is a 51/49 split across men/women in the initial survey, but this is skewed slightly in favour of men in our sample, likely driven by selection on occupation as discussed above. In our models we condition on measures of prior achievement, based on a point score from external tests taken at the end of primary school (Key Stage 2 – KS2) which is supplied through the matched administrative data. Our sample performed slightly better in the KS2 tests relative to the initial sample. Finally we can also condition on measures of ethnic group in our models, which is reported in wave 3 (age 16). 27.1% of our sample report being in in a Black, Asian, or Minority Ethnic group, relative to 32% of the original survey respondents. When we split our sample by attainment quintiles, we do this based on their ranking in the original national distribution of attainment, although there are relatively similar proportions of high and low attainers in our final sample after weighting.

We begin by showing flexible graphical representations of inequalities in mismatch by plotting the survey respondents' achievement decile against the average occupation quality of all individuals in that decile. If people are perfectly matched to occupations, then individuals in the bottom decile would, on average, work in occupations at the 1st decile of the occupation ranking, while individuals in the seventh decile would work in occupations at the 7th decile of occupations, illustrated by the 45 degree line on the graphs. We also show SES and gender inequalities in occupational match conditional on prior achievement and demographics, estimated from the following regression:

$$M_{ia} = \beta_0 + \beta_1 SES_i + \gamma female_i + \pi Key\ Stage\ 2_i + \rho ethnicity_i + \varepsilon_i, \Delta a, \qquad (1)$$

where M_i is our measure of education-based and earnings-based match, $\widehat{\beta_1}$ is our estimated SES gap in match, and $\hat{\gamma}$ is our estimates gender gap in match, conditional on prior achievement and ethnicity. Given that match is defined using individual rankings, there are limits at the top and the bottom of the distribution making it impossible for the lowest-ranked student to undermatch, and the highest-ranked student to overmatch. We therefore estimate the models separately across deciles and quintiles of achievement, a. All results are weighted using the wave 8 final weights.

To explore the role of market failures and preferences in driving SES and gender gaps in mismatch, we add potential mediators, Med_i, separately to the model to assess whether they lead to a reduction in our estimates $\widehat{\beta_1}$ and $\hat{\gamma}$.

$$M_{ia} = \beta_0 + \beta_1 SES_i + \gamma female_i + \tau Med_i + \pi Key\ Stage\ 2_i + \rho ethnicity_i + \varepsilon_i, \Delta a, \quad (2)$$

We consider the role of educational pathways, including both post-compulsory schooling outcomes (Key Stage 5 points at age 18, university institution[5] and subject choice[6]), and post-compulsory schooling decisions (staying on post-16, staying on post-18, and whether employed or not post-18), first separately and them combined. This allows us to assess whether any SES or gender gaps in occupational match are purely driven by differential decisions about education and labour market experiences after age 16.

To examine the existence of preferences with regards to type of employment, we consider the role of industry, using the 3-digit Standard Industry Classification (SIC) of the business establishment they are employed by at age 25. While this is not accounting for individuals' non-pecuniary preferences over occupations, it does account for preferences over industries with varied work schedules, and other non-pecuniary benefits, including location. For example, an accountant, software designer, or administrative assistant could work across multiple different industries with very different experiences.

It is well established in the literature that young people are more likely to work in the same occupations as their parents, which can represent both preferences for particular types of jobs, and market failures in terms of information constraints about other occupations (Corak and

[5] Oxbridge, Russell Group, and other institution

[6] 18 JACS Principal Subject Codes including Medicine & Dentistry, Subjects Allied to Medicine, Biological Sciences, Veterinary sciences, Agriculture and related subjects, Physical sciences, Mathematical sciences, Computer science, Engineering and Technology, Architecture, building and planning, Social studies, Law, Business and administrative studies, Mass communications & documentation, Languages, Historical and philosophical studies, Creative arts & design, Education

Piraino, 2011). We observe the occupation of the main respondent parent at age 14, which we can account for in our model. The detailed survey data also offers insights into the respondents' preferences in terms of their reasons for wanting to work in certain occupations at age 20 (wave 7), including to help people, to be paid well, to be the boss, to have a non-routine job, to have good promotion chances, and to have regular working hours.

Finally, we observe some proxies for the role of market failures through measures of local labour market conditions (Government Office Region at age 25), questions on academic self-concept during school (Hansen and Henderson, 2019), and the use of information, advice, and guidance (IAG), and the most useful types of IAG used at age 20 (wave 7) when considering career options.

3. Results

Figure 3 shows the distribution of education- and earnings- based occupational match for our sample. The distributions are broadly symmetrical, although earnings-based match is more right-skewed than education-based match, suggesting more undermatch in terms of earnings-based rankings of occupation. Using a binary definition of mismatch of +/- 20 percentiles, 26% of the sample are undermatched and 20% overmatched in our education-based measure, while 31% are undermatched and 16% overmatched in our earnings-based measure. This illustrates the extent of inefficiencies in the labour market in terms of the match between individual quality and occupation quality across both dimensions of match, with almost half of the sample mismatched by over 20 percentiles into lower/higher quality occupations than their own quality would suggest.

Figure 4 shows how this mismatch presents in terms of socio-economic differences across the distribution of individual achievement. If individuals were perfectly matched to their occupations, the lines would sit at 45 degrees. As we can see from Figure 4, individuals from both high and low SES backgrounds mismatch into occupations, with flatter lines across the distribution illustrating more overmatch among low achievers and more undermatch among high achievers on average for both education-based and earnings-based match. Crucially, at every point in the distribution of individual achievement, people from low SES backgrounds work in lower quality occupations than people form high SES backgrounds, with particularly pronounced differences at the 9th decile of achievement, and more convergence for median achievement.

In Figure 5, we plot average occupation quality across deciles of individual achievement for women and men separately. Here we see an interesting difference, with women working in higher ranked occupations than men, in terms of education-based ranks, but for earnings-based ranks, women are working in lower ranked occupations across the entire distribution, and in particular for those at the bottom of the achievement distribution. This is consistent with findings from the higher education mismatch literature where women attend similarly selective courses as men in terms of academic achievement, but are undermatching in terms of earnings-based course rankings (Campbell et al., 2022).

Figures 6 and 7 (and Appendix Table A1) present estimates of SES and gender match gaps conditional on demographics and prior achievement at age 11 across the distribution of individual achievement. We regress the match index (education-based in the top panel and earnings-based in the bottom panel) on measures of SES, gender, and background characteristics as described in equation (1) in the previous section. Focusing first on SES in Figure 6, we can see that there no SES gap in occupation mismatch at the bottom of the achievement distribution for both measures of match, but for the second quintile of achievement individuals from low SES backgrounds are undermatching 7-8 percentiles more than individuals from high SES backgrounds, with the same achievement levels. This SES gap is slightly smaller for the third achievement quintile (5 percentiles) and disappears for the fourth achievement quintile, but comes back strongly at the top of the achievement distribution where individuals from low SES backgrounds work in occupations that are ranked 11 percentiles lower than individuals from high SES backgrounds. There are strikingly similar patterns in the SES gaps in mismatch across earnings and education-based measures of occupational mismatch, which is striking given that they are measuring different concepts.

By contrast, as seen in Figure 5, the results for gender gaps in match differ based on our match metric. Figure 7 shows clearly that for education-based occupation quality ranking, women typically work in occupations that are higher ranked in terms of the average level of highest education of workers in those jobs for the majority of the achievement distribution. At the second and third quintile of achievement, women on average work in occupations ranked 5-6 percentiles higher than men in terms of education-based quality. In the top two achievement quintiles, there is a 3 percentile gap in favour of women working in higher ranked occupations.

For earnings-based quality rankings however, women systematically work in lower ranked jobs across the achievement distribution, ranging from working in occupations ranked 16 percentiles lower than men for the lowest achievement quintiles, 10 percentiles lower in the second quintile, 5 percentiles lower in quintiles 3 and 4, and 3 percentiles lower at the top of the achievement distribution. This suggests that women work in occupations that are more demanding in terms of education levels required, relative to men, but pay significantly less.

In Appendix Table A2 we check that our results hold if we use an alternative measure of parental SES, namely parental education. We find reassuringly similar results when using measures of parental education, with no SES gaps at the bottom of the achievement distribution but large SES gaps among high achievers, with individuals from low SES backgrounds (low educated parents) working in occupations ranked 12 percentiles lower than individuals from high SES backgrounds (highly educated parents). The gender patterns remain stable for this alternative specification with positive gender gaps in terms of education-based occupation rankings among high achieving women, and a large penalty for women in terms of earnings-based occupation rankings among low achievers.

4. Mechanisms

Given our rich survey data, we can explore why we find large SES and gender gaps in occupational mismatch across the distribution of achievement. Here we focus on earnings-based measures of match for reasons of brevity but our main findings for SES are very similar for education-based measures of match.

Table 4 presents a description of our mediator variables across SES and gender. As is expected, individuals from low SES backgrounds score lower in terms of Key Stage 5 points, are more likely to not be in full time education post-16, and are less likely to go to university than their high SES counterparts. Education pathways beyond compulsory schooling could therefore account for our SES gaps in occupational mismatch based on definition of individual quality as their position in the ranking of GCSE scores at age 16. Similarly, women score higher than men in terms of Key Stage 5 points scores, and are more likely to go to university than men. These pathways could also therefore explain some of the gender gap in occupational mismatch.

Figure 8 plots SES gaps in earnings-based mismatch for low and high achievers, presenting first the baseline estimates from Figure 6, before considering the role of post-16 and post-18 education separately, and then all education pathways together for the bottom achievement quintile (left) and the top achievement quintile (right). There is no SES gap in occupational mismatch among low achievers, but the SES gap among high achievers is reduced by around 21-26% (2-3 percentiles) when accounting for post-16 and post-18 education separately, and 33% (4 percentiles) when accounting for all education pathways post-16. This suggests that while post-16 education pathways do account for some fraction of the occupational mismatch observed at age 25, the majority is not working through post-compulsory experiences.

We can also consider the importance of parental occupation in accounting for our SES and gender gaps. Individuals from low SES backgrounds have parents who work in occupations with lower SOC codes than their high SES peers, but parental occupation is similarly distributed by gender. In Figure 8 we present estimates controlling for parental occupation compared to our baseline SES estimates. There is a slight reduction in the SES gaps in occupational mismatch, particularly among high achievers, when conditioning on parental occupations (15% or 1.7 percentiles), suggesting that a small part of the difference in occupational mismatch observed, namely lower SES individuals working in lower quality occupations, is driven by their parent's occupations. As mentioned, this could be capturing preferences or market failures through information constraints based on parental occupations.

The wealth of survey data available in Next Steps allows us to examine the role of motivations for occupation choices, academic self-concept, and the use of information, advice, and guidance (IAG) in accounting for SES and gender gaps in occupational mismatch. Table 4 shows some slight differences in motivation for different jobs by SES, with individuals from low SES backgrounds more likely to want a job that pays well, with regular hours, compared to those from high SES backgrounds who prefer to be their own boss and have non-routine jobs. Men prefer jobs with good promotion chances and less routine, while women prefer to be their own boss and regular hours. Academic self-concept varies by SES with low SES individuals reporting lower levels of academic self-concept, relative to their high SES counterparts. Low SES individuals are also less likely to use IAG, and in particular less likely to use friends and relatives to find out about jobs, compared to high SES counterparts. Women and men in our sample report very similar levels of academic self-concept, with women more likely to use IAG and in particular use friends and relatives to find out about occupations, relative to men.

Figure 8 shows that accounting for motivation for occupation choice, academic self-concept and information, advice and guidance do not change our estimates of SES gaps in occupational mismatch among high or low achievers, suggesting they are not major reasons as to why low SES individuals work in lower ranked occupations than high SES individuals.

We can also consider the importance of industry choice and local region at age 25 in accounting for our SES and gender gaps. Table 4 illustrates that low SES individual's work in slightly different industries to high SES individuals, but there is more difference between industries for women and men. By contrast, individuals from low SES backgrounds are also more likely to be found in the North while high SES young people are more likely to be live in the South of England, but the distribution of region looks more similar by gender. In Figure 8 we present estimates controlling for industry of employment and region at age 25 compared to our baseline SES estimates. SES gaps among both low and high achievers are reduced by around 3 percentiles when comparing individuals that work in the same industries, suggesting that industry selection is accounting for some of the SES gap in occupational mismatch we observe. There is also a slight reduction in the SES gaps in occupational mismatch, particularly among high achievers, when conditioning on region (27% of 3 percentiles), suggesting that a small part of the difference in occupational mismatch observed, namely lower SES individuals working in lower quality occupations, is driven by their spread across the country.

Figure 9 plots gender gaps in mismatch for low achievers (left) and high achievers (right), first presenting the baseline estimates from Figure 7, before adding mechanisms. For low achievers, the large gender penalty is unmoved by the inclusion of education pathway measures, suggesting that the mismatch that arises due to women working in lower quality occupations than men in terms of their GCSE rankings is not driven by the educational choices they make beyond compulsory schooling. The finding also holds among high achievers.

The gender gap in occupational mismatch among low achievers is significantly reduced once we account for industry of employment, with the gender difference reducing by 76% (12 percentiles) when comparing women and men working in the same industries. This indicates that the vast majority of the large occupational mismatch gender gap among low achievers is working through selection into different industries of women relative to men. This could be indicating an important role for preferences in terms of occupation differences between women and men. Yet, the high achieving gap remains relatively stable suggesting industry selection is not accounting for differences at the top of the achievement distribution. In addition, a significant gender gap in occupational mismatch remains, with women working in lower ranked occupations than men in terms of hourly pay, among both low and high achievers after comparing individuals who work in the same industry.

When we account for parental occupation, there is very little difference in the gender gap, suggesting that women working in lower ranked occupations than men is not being driven by differences in parental occupation between women and men. Similarly, the destination region of the young person at 25 is making no difference to the gender gap, suggesting this gap occurs within regions rather than between. Academic self-concept and IAG also do very little to account for gender gaps in occupational mismatch. But motivations for occupational choice can account for a small portion of the gender gap among low achievers (14% or 2 percentiles). This suggests that part of the explanation for women working in lower ranked occupations than men at the bottom of the achievement distribution is different motivations for their occupation choices.

5. Conclusions

While previous research has documented penalties to overeducation, and inequalities in access to top jobs, even after accounting for educational qualifications, little is known about how individuals match to occupations across the distribution of achievement, and how this varies by key demographics. Our new measures of individual to occupation mismatch show large amounts of mismatch across the distribution of individual achievement, with almost half of all early labour market workers being mismatched in terms of education-based and earnings-based match.

We also document large inequalities in occupational mismatch, with individuals from low SES backgrounds working in lower quality occupations across the distribution of achievement for both measures of match. While these penalties are smaller among low achievers, they are sizable at both the middle and the top of the achievement distribution. We show that post-16 educational pathways can only account for a small portion of this SES gap, with the majority of the SES gap remaining when comparing individuals with very similar post-16 education profiles. This is consistent with previous literature on inequalities in access to top jobs after accounting for educational differences (Macmillan et al., 2015). Similarly, industry choice and parental occupation can only account for a small part of SES gaps in occupational mismatch, while measures of motivation for occupation choice, academic self-concept, and IAG have no explanatory power for understanding why low SES individuals work in lower ranked occupations than we might expect, given their achievement.

There are interesting differences in gender gaps in occupational mismatch. Women typically work in occupations that are ranked slightly higher than men in terms of average education levels of workers in those occupations. Yet when we look at earnings-based occupational mismatch, we see that women work in occupations that are significantly lower ranked than men, particularly among low achievers. We find that industry selection can account for a large proportion of this gender difference, but women still work in lower ranked occupations than men in terms of hourly pay, even when comparing people working in the same industries. Further, there is evidence that motivations for occupational choice may be driving a small part of the gender gap among low achievers, but education pathways, parental occupation, academic self-concept, and IAG use are not major drivers of differences in occupation mismatch between women and men.

The importance of industry choice does suggest some role for preferences in our findings of occupational mismatch, yet both SES and gender gaps in occupational mismatch remain even after accounting for industry of work. The importance of market failures, in the form of information constraints through parental occupations, academic self-concept, and IAG are more muted. While the re-ranking of individuals based on later education choices post-16 can account for a small part of the occupational mismatch observed, this is clearly not the main reason for the SES and gender inequalities in mismatch observed. Future research should consider other channels through which low SES individuals undermatch into occupations.

References

Baert, S., Cockx, B., & Verhaest, D. (2013). “Overeducation at the start of the career: Stepping stone or trap?” *Labour Economics* 25 (December). European Association of Labour Economists 24th Annual Conference, Bonn, Germany, 20–22 September 2012: 123–40. doi:10.1016/j.labeco.2013.04.013.

Battu, H., Belfield, C. R., & Sloane, P. J. (2000). “How well can we measure graduate over-education and its effects?” *National Institute Economic Review, 171*(1), 82–93. doi:10.1177/002795010017100107.

Bauer, T. K. (2002). “Education mismatch and wages: A panel Analysis”. *Economics of Education Review, 21*(3), 221–229. doi:10.1016/S0272-7757(01)00004-8.

Campbell, S., Macmillan, L., Murphy, R., Wyness, G. (2022) “Matching in the Dark? Inequalities in Student to Degree match” *Journal of Labor Economics* Vol 40(4).

Crawford, C., and Vignoles, A. (2014) “Heterogeneity in graduate earnings by socio-economic background” *IFS working paper* 201430.

Corak, M., and Piraino, P., (2011) “The Intergenerational Transmission of Employers” *Journal of Labor Economics* 29(1) -

Dickson, M., Donnelly, M., Kameshwara, L., Lazetic, P. (2022) “The scarring effect of graduate underemployment: evidence from the UK” mimeo University of Bath.

Elias, P., & Purcell, K. (2004). Is Mass Higher Education Working? Evidence from the Labour Market Experiences of Recent Graduates. *National Institute Economic Review, 190*, 60-74. doi:10.1177/002795010419000107

Flisi, S., Goglio, V., Meroni, E.C., Rodrigues, M., Vera-Toscano, E. (2017). “Measuring Occupational Mismatch: Overeducation and Overskill in Europe—Evidence from PIAAC”. *Social Indicators Research* 131, 1211–1249

Green, F., and Henseke, G. (2016a) “Should governments of OECD countries worry about graduate under-employment?” *Oxford Review of Economic Policy* Vol. 32(4).

Green, F., and Henseke, G. (2016b). “The changing graduate labour market: analysis using a new indicator of graduate jobs”. *IZA Journal of Labor Policy*,5(1): 14.

Green, F., & Zhu, Y. (2010). “Overqualification, job dissatisfaction, and increasing dispersion in the returns to graduate education”. *Oxford Economic Papers, 62*(4), 740–763. doi:10.1093/oep/gpq002.

Guvenen, F., Kuruscu, B., Tanaka, S., and Wiczer, D. (2020). "Multidimensional Skill Mismatch." *American Economic Journal: Macroeconomics*, 12 (1): 210-44.

Hansen, K., and Henderson, M. (2019) “Does academic self-concept drive academic achievement?” *Oxford Review of Education* 45 (5). 657-672.

McGuinness, S. (2006). “Overeducation in the labour market”. *Journal of Economic Surveys, 20*(3), 387–418. doi:10.1111/j.0950-0804.2006.00284.x.

Macmillan, L., Tyler, C., Vignoles, A., (2015) "Who Gets the Top Jobs? The Role of Family Background and Networks in Recent Graduates' Access to High-status Professions" *Journal of Social Policy* Vol 44(3).

Mavromaras, K., Mahuteau, S., Sloane, P., and Wei, Z. (2013). "The effect of overskilling dynamics on wages". *Education Economics*, 21(3): 281-303.

Quinn, M., and Rubb, S. (2006) "Mexico's labor market: The importance of education-occupation matching on wages and productivity in developing countries" *Economics of Education Review* 25 (2). 147-156.

Office for National Statistics, Social Survey Division, Northern Ireland Statistics and Research Agency, Central Survey Unit. (2019). *Quarterly Labour Force Survey, January - March, 2014*. [data collection]. *9th Edition.* UK Data Service. SN: 7501, http://doi.org/10.5255/UKDA-SN-7501-9

Northern Ireland Statistics and Research Agency, Central Survey Unit, Office for National Statistics, Social Survey Division. (2019). *Quarterly Labour Force Survey, April - June, 2014*. [data collection]. *6th Edition.* UK Data Service. SN: 7557, http://doi.org/10.5255/UKDA-SN-7557-6

Office for National Statistics, Social Survey Division, Northern Ireland Statistics and Research Agency, Central Survey Unit. (2019). *Quarterly Labour Force Survey, July - September, 2014*. [data collection]. *4th Edition.* UK Data Service. SN: 7570, http://doi.org/10.5255/UKDA-SN-7570-4

Office for National Statistics, Social Survey Division, Northern Ireland Statistics and Research Agency, Central Survey Unit. (2019). *Quarterly Labour Force Survey, October - December, 2014*. [data collection]. *4th Edition.* UK Data Service. SN: 7664, http://doi.org/10.5255/UKDA-SN-7664-4

Northern Ireland Statistics and Research Agency, Central Survey Unit, Office for National Statistics, Social Survey Division. (2019). *Quarterly Labour Force Survey, January - March, 2015*. [data collection]. *5th Edition.* UK Data Service. SN: 7725, http://doi.org/10.5255/UKDA-SN-7725-5

Office for National Statistics, Social Survey Division, Northern Ireland Statistics and Research Agency, Central Survey Unit. (2019). *Quarterly Labour Force Survey, April - June, 2015*. [data collection]. *5th Edition.* UK Data Service. SN: 7781, http://doi.org/10.5255/UKDA-SN-7781-5

Northern Ireland Statistics and Research Agency, Central Survey Unit, Office for National Statistics, Social Survey Division. (2019). *Quarterly Labour Force Survey, July - September, 2015*. [data collection]. *6th Edition.* UK Data Service. SN: 7842, http://doi.org/10.5255/UKDA-SN-7842-6

Office for National Statistics, Social Survey Division, Northern Ireland Statistics and Research Agency, Central Survey Unit. (2019). *Quarterly Labour Force Survey, October - December, 2015*. [data collection]. *5th Edition.* UK Data Service. SN: 7902, http://doi.org/10.5255/UKDA-SN-7902-5

Office for National Statistics, Social Survey Division, Northern Ireland Statistics and Research Agency, Central Survey Unit. (2021). *Quarterly Labour Force Survey, January - March, 2016*. [data collection]. *6th Edition.* UK Data Service. SN: 7985, http://doi.org/10.5255/UKDA-SN-7985-6

Office for National Statistics, Social Survey Division, Northern Ireland Statistics and Research Agency, Central Survey Unit. (2019). *Quarterly Labour Force Survey, April - June, 2016*. [data collection]. *3rd Edition.* UK Data Service. SN: 8039, http://doi.org/10.5255/UKDA-SN-8039-3

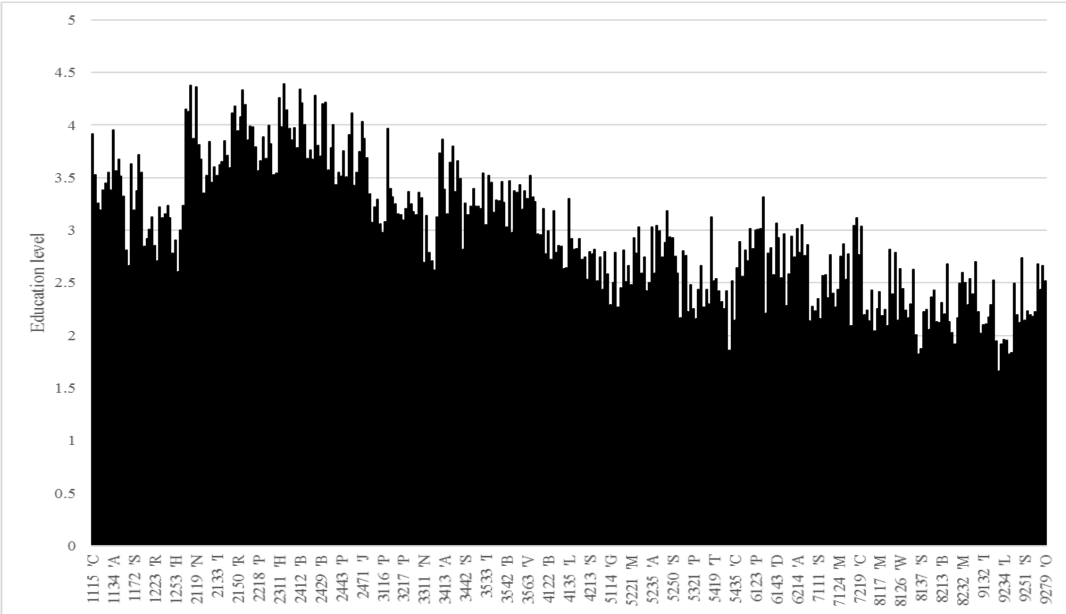
Northern Ireland Statistics and Research Agency, Central Survey Unit, Office for National Statistics, Social Survey Division. (2019). *Quarterly Labour Force Survey, July - September, 2016*. [data collection]. *5th Edition.* UK Data Service. SN: 8104, http://doi.org/10.5255/UKDA-SN-8104-4

Office for National Statistics, Social Survey Division, Northern Ireland Statistics and Research Agency, Central Survey Unit. (2019). *Quarterly Labour Force Survey, October - December, 2016*. [data collection]. *3rd Edition.* UK Data Service. SN: 8145, http://doi.org/10.5255/UKDA-SN-8145-3

University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2021). *Next Steps: Sweeps 1-8, 2004-2016: Secure Access*. [data collection]. *2nd Edition.* UK Data Service. SN: 8656, http://doi.org/10.5255/UKDA-SN-8656-2

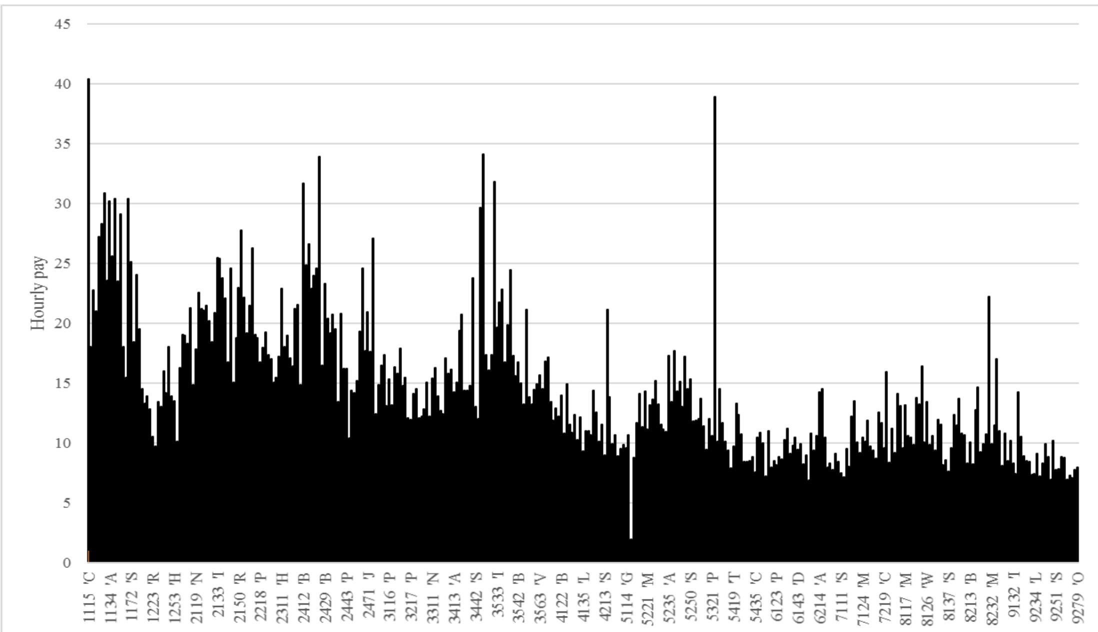
Charts

Figure 1 Distribution of education by occupation



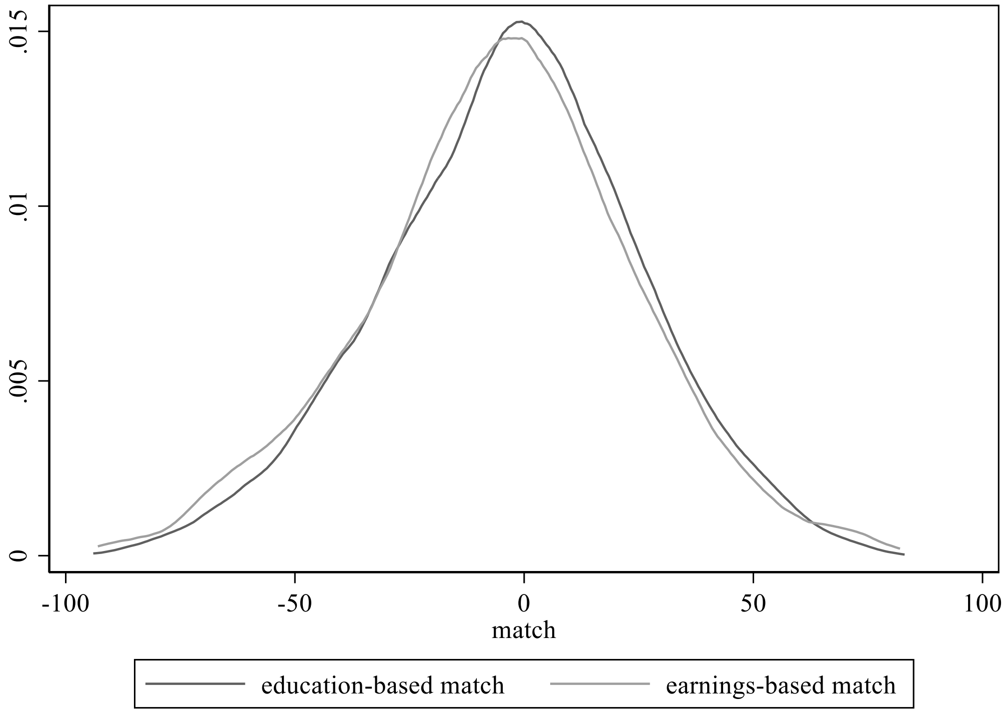
N=433977. Secondary education teachers (2314), Physical scientists (2113), and Natural science professionals (2119) have the highest education. Street cleaners (9232), Tyre, exhaust, and windscreen fitters (8135), and vehicle valets and cleaners 99236) have the lowest education levels.

Figure 2 Distribution of earnings by occupation



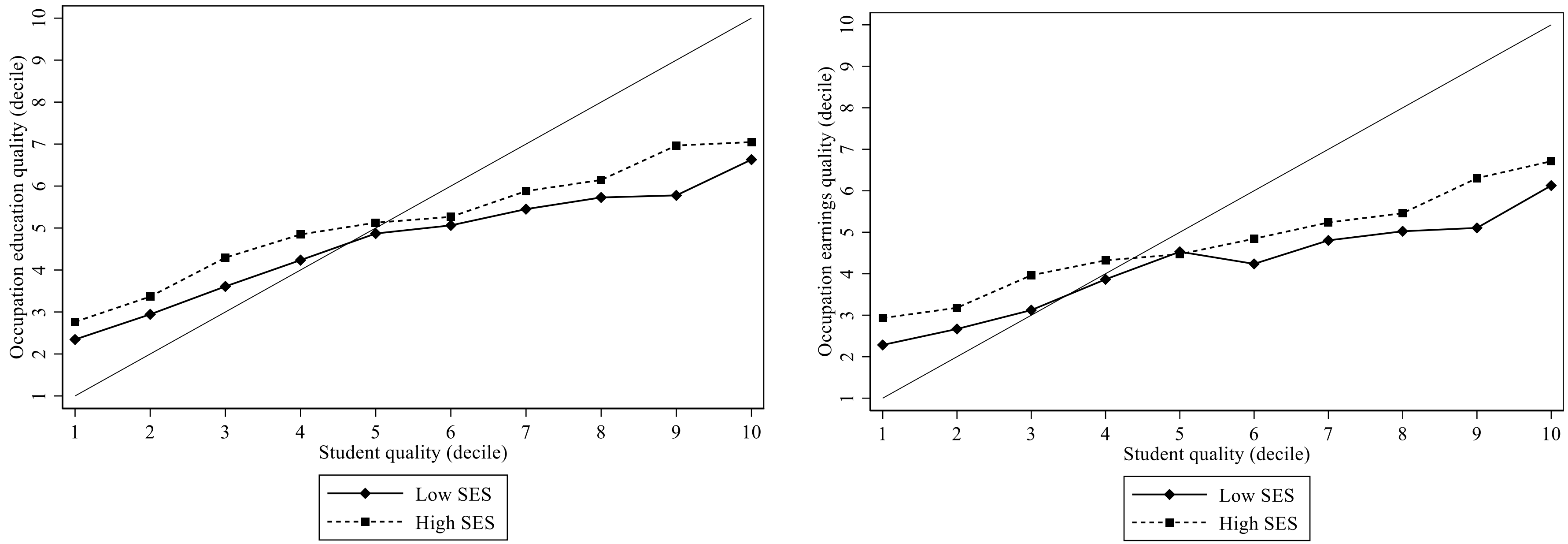
N =105352. CEOs (1115), aircraft pilots (3512), and floorers and wall tilers (5322) have the highest hourly pay. Care escorts (6147), School midday and crossing patrol occupations (9244), and kitchen and catering assistants (9272) have the lowest pay.

Figure 3 Distribution of education-based and earnings-based occupational mismatch



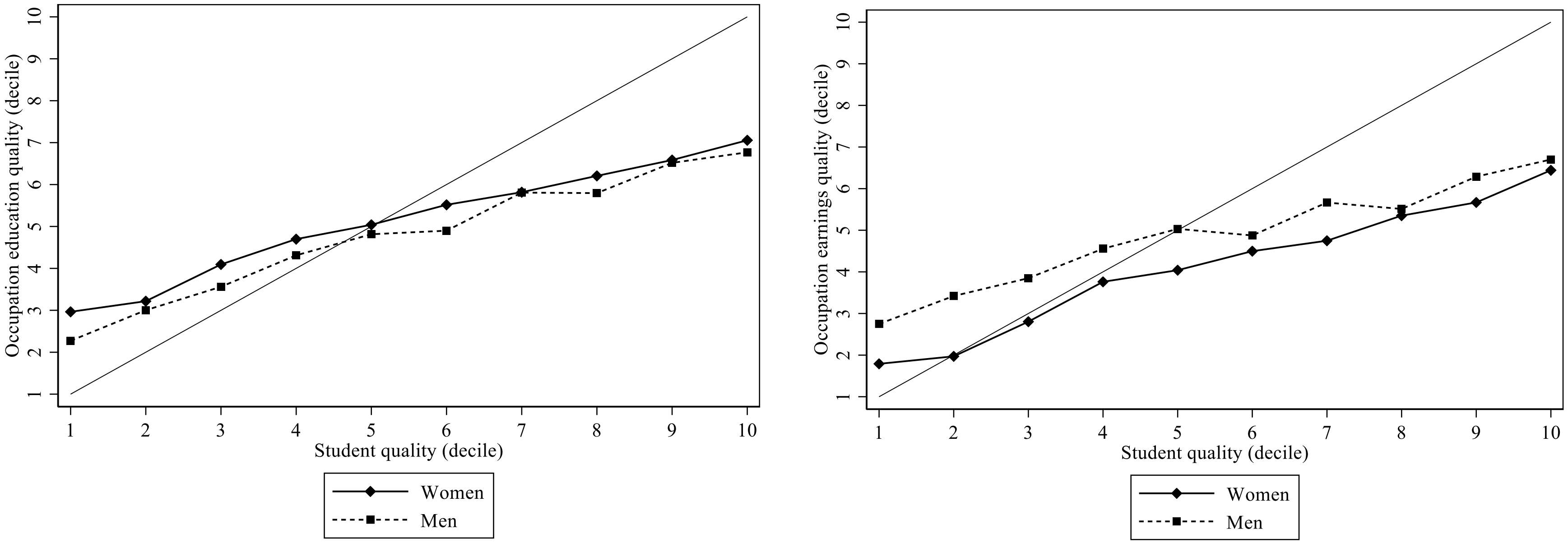
Notes: Source, Next Steps. Sample N=4744. Our sample is defined as survey respondents with both a GCSE total points score from the matched NPD data and a 4-digit occupational SOC code at age 25. Wave 8 final weights applied.

Figure 4 SES match by student achievement, for education-based (left) and earnings-based (right) match



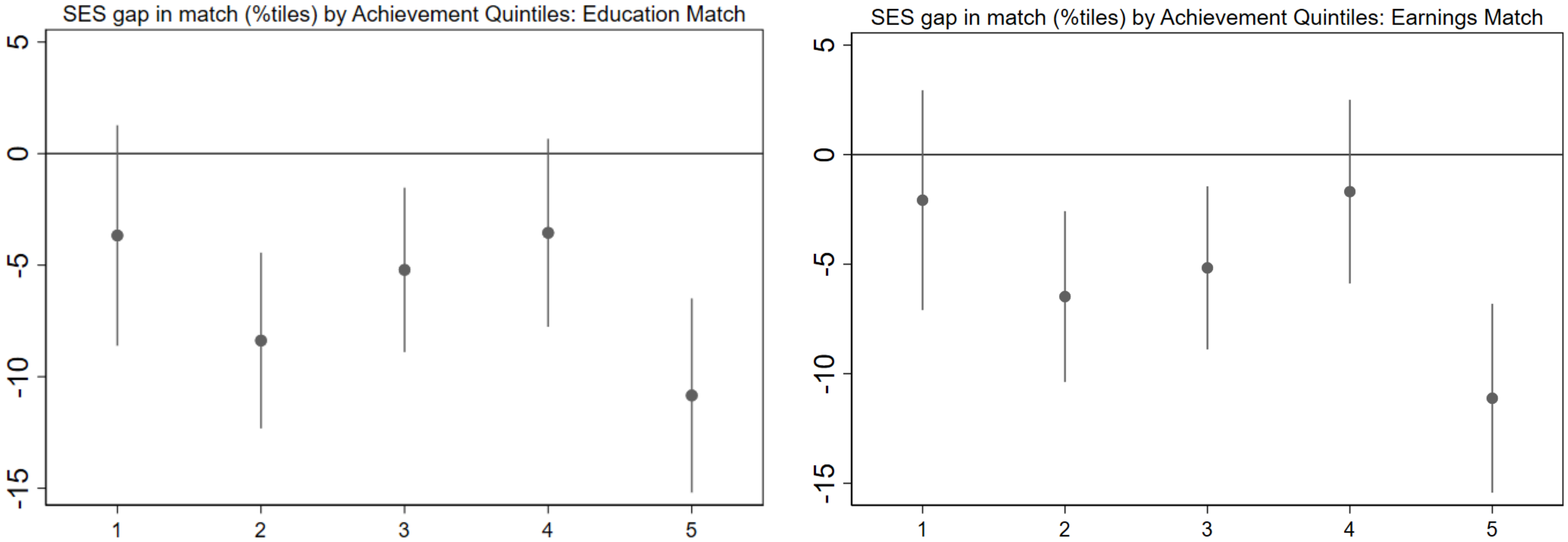
Notes: Source, Next Steps. Sample N=1285 low SES, and N=1967 high SES. The 45 degree line represents perfect matching throughout the achievement distribution. Student quality defined by their age 16 GCSE points score in national distribution from NPD. Occupation quality defined by average highest education qualification of workers age 25-60 in national distribution from LFS. Quality measures converted to percentiles from which deciles are obtained. Our sample is defined as survey respondents with both a GCSE total points score from the matched NPD data and a 4-digit occupational SOC code at age 25. Wave 8 final weights applied.

Figure 5 Gender match by student achievement, for education-based (left) and earnings-based (right) match



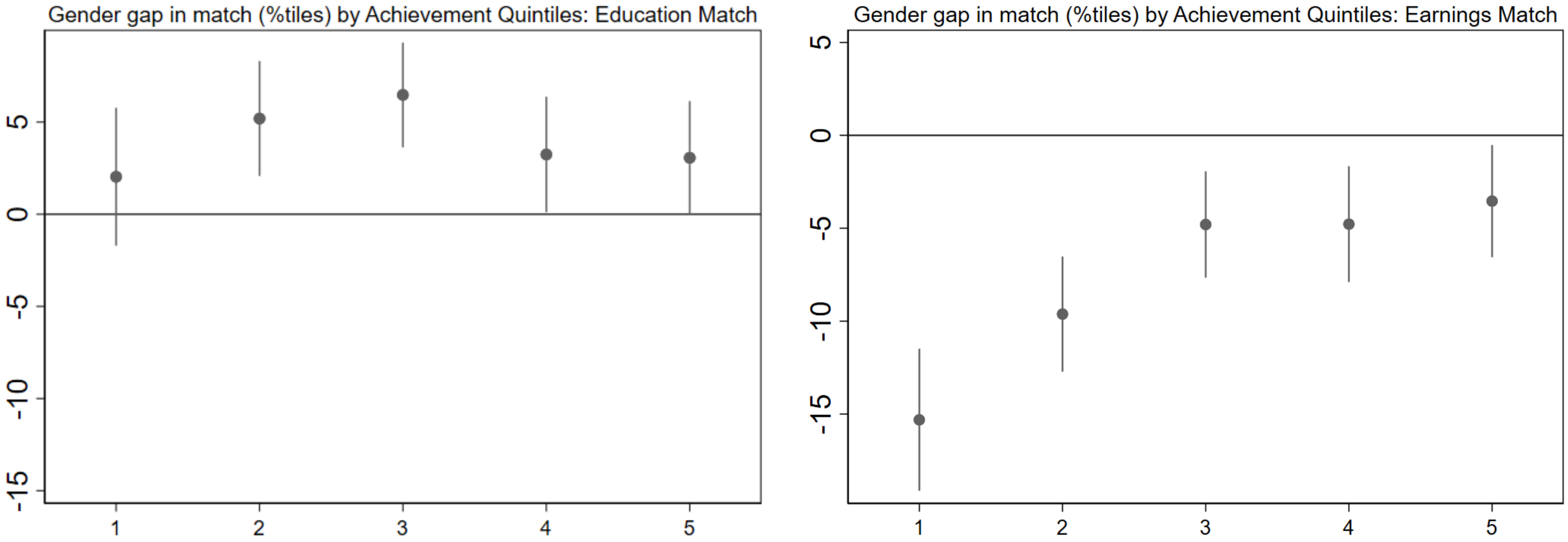
Notes: Source, Next Steps. Sample N=2548 women, N=2196 men. The 45 degree line represents perfect matching throughout the achievement distribution. Student quality defined by their age 16 GCSE points score in national distribution from NPD. Occupation quality defined by average highest education qualification of workers age 25-60 in national distribution from LFS. Quality measures converted to percentiles from which deciles are obtained. Our sample is defined as survey respondents with both a GCSE total points score from the matched NPD data and a 4-digit occupational SOC code at age 25. Wave 8 final weights applied.

Figure 6: SES gaps in mismatch in education-based and earnings-based match across quintiles of achievement



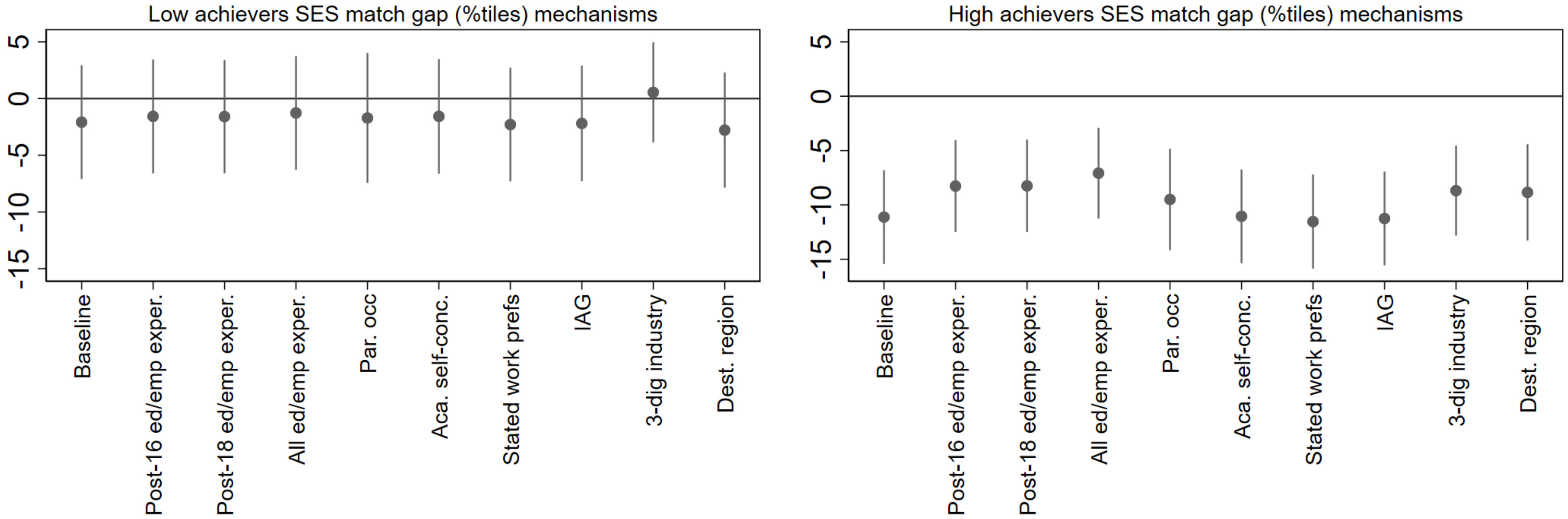
Notes: Source, Next Steps. Samples N=518, 896, 1215, 1086, 1029 for quintiles 1-5. Our sample is defined as survey respondents with both a GCSE total points score from the matched NPD data and a 4-digit occupational SOC code at age 25. Wave 8 final weights applied. Controls include Key Stage 2 points score (age 11) and a categorical measure of ethnicity. SES gaps show mismatch for low SES respondents relative to high SES respondents.

Figure 7: Gender gaps in mismatch in education-based and earnings-based match across quintiles of achievement



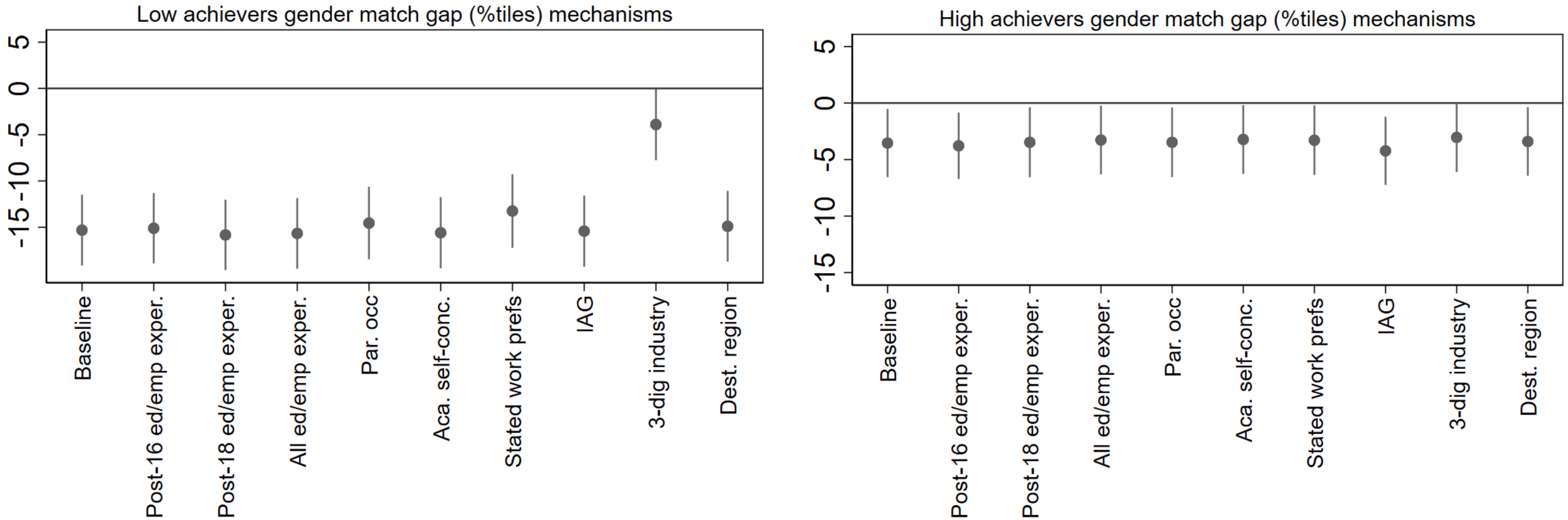
Notes: Source, Next Steps. Samples N=518, 896, 1215, 1086, 1029 for quintiles 1-5. Our sample is defined as survey respondents with both a GCSE total points score from the matched NPD data and a 4-digit occupational SOC code at age 25. Wave 8 final weights applied. Controls include Key Stage 2 points score (age 11) and a categorical measure of ethnicity. Gender gaps show mismatch for women relative to men.

Figure 8: SES gaps in mismatch (earnings-based), conditional on a range of mediators for low and high achieving individuals



Notes: Source, Next Steps. Samples N=518, 1209. Our sample is defined as survey respondents with both a GCSE total points score from the matched NPD data and a 4-digit occupational SOC code at age 25. Wave 8 final weights applied. Low attainers are those in the bottom 20 percentiles of GCSE attainment at age 16. High attainers are those in the top 20 percentiles of GCSE attainment at age 16. Controls include Key Stage 2 points score (age 11) and a categorical measure of ethnicity. Age 16 to age 18 activity: KS5 points scores, not in full time ed age 16-18. Age 18 to 21 activity: In university, working, not in education, employment or training, university institution attended (Oxbridge, RG, other), subject studied (18 JACS principal subject codes). 3 digit SIC: 3-digit industry (SIC code) of the respondent at age 25. Par 2-dig SOC: 2-digit occupation (SOC code) of the main parent in wave 4 (age 16). Destination region: Government office region of the respondent at age 25. Motivation for occupation choice at 20: Statements about what desire from job, including to help other people, a job that pays well, to be my own boss, to have an interesting / non-routine job, to have a chance of promotion, to have regular hours. Academic self-concept: quartiles of academic self-concept score. Information, advice, and guidance: Most useful source of IAG in the past 12 months at age 20, including no IAG used, friends and relatives, teachers and lecturers, connexions, Direct.gov, National Apprenticeship Service, Jobcentre plus advisor, Professional in field of interest, Careers advisors and student support services. Each set of mediators added separately, with the exception of model 4, which is model 2 and model 3 combined. SES gaps show mismatch for low SES respondents relative to high SES respondents.

Figure 9: Gender gaps in mismatch (earnings-based), conditional on a range of mediators for low and high achieving individuals



Notes: Source, Next Steps. Samples N=518, 1209. Our sample is defined as survey respondents with both a GCSE total points score from the matched NPD data and a 4-digit occupational SOC code at age 25. Wave 8 final weights applied. Low attainers are those in the bottom 20 percentiles of GCSE attainment at age 16. High attainers are those in the top 20 percentiles of GCSE attainment at age 16. Controls include Key Stage 2 points score (age 11) and a categorical measure of ethnicity. Age 16 to age 18 activity: KS5 points scores, not in full time ed age 16-18. Age 18 to 21 activity: In university, working, not in education, employment or training, university institution attended (Oxbridge, RG, other), subject studied (18 JACS principal subject codes). 3 digit SIC: 3-digit industry (SIC code) of the respondent at age 25. Par 2-dig SOC: 2-digit occupation (SOC code) of the main parent in wave 4 (age 16). Destination region: Government office region of the respondent at age 25. Motivation for occupation choice at 20: Statements about what desire from job, including to help other people, a job that pays well, to be my own boss, to have an interesting / non-routine job, to have a chance of promotion, to have regular hours. Academic self-concept: quartiles of academic self-concept score. Information, advice, and guidance: Most useful source of IAG in the past 12 months at age 20, including no IAG used, friends and relatives, teachers and lecturers, connexions, Direct.gov, National Apprenticeship Service, Jobcentre plus advisor, Professional in field of interest, Careers advisors and student support services. Each set of mediators added separately, with the exception of model 4, which is model 2 and model 3 combined. Gender gaps show mismatch for women relative to men.

Tables

Table 1: Key characteristics for full cohort compared to occupation mismatch sample (selected on 4-digit occupation soc code)

	Full cohort		Sample	
	Mean	N	Mean	N
Low SES	33.6	13,902	29.6	4,744
Medium SES	31.3	13,902	32.9	4,744
High SES	34.3	13,902	37.4	4,744
Women	50.9	15,431	53.5	4,744
Men	49.1	15,431	46.5	4,744
KS2 pts	80.8	14,151	82.5	4,744
BAME	32.3	11,783	27.1	4,744
GCSE ptile	48.5	11,874	48.8	4,744
High attainer	19.3	11,874	19.6	4,744
Low attainer	20.0	11,874	19.0	4,744

Notes: Source, Next Steps. Samples reported in columns 2 and 4. Our sample is defined as survey respondents with both a GCSE total points score from the matched NPD data and a 4-digit occupational SOC code at age 25. Wave 8 final weights applied.

Table 2: Age 25 (wave 8) main activity for full cohort compared to occupation mismatch sample (selected on 4-digit occupation soc code)

	Full cohort	Sample	Full cohort	Sample	Full cohort	Sample
	All	All	Men	Men	Women	Women
Employee	74.6	91.3	74.9	87.6	74.7	94.4
Self-employed	6.5	7.8	9.8	11.3	4.0	4.9
Unemployed	5.7	0.0	6.4	0.0	4.9	0.0
Education	5.0	0.0	4.7	0.0	5.1	0.0
Looking after family	4.7	0.0	0.4	0.0	8.2	0.0
Other	3.5	0.9	3.8	1.1	3.1	0.7
Total	100	100	100	100	100	
N	7,707	4,744	3,321	2,196	4,153	2,548

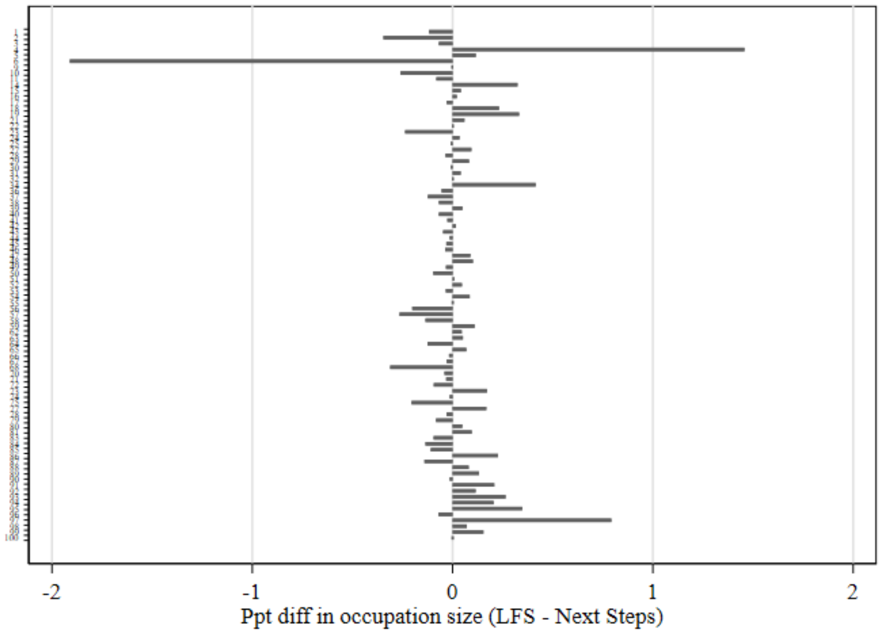
Notes: Source, Next Steps. Samples reported in final row. Our sample is defined as survey respondents with both a GCSE total points score from the matched NPD data and a 4-digit occupational SOC code at age 25. Wave 8 final weights applied.

Table 3: Potential occupational mismatch mediators, by SES (parental NS-SEC) and gender

	SES		Gender	
	Low	High	Women	Men
KS5 points	267	523	439	345
Not in FT ed 18	55%	38%	54%	49%
Uni	29%	60%	49%	39%
Work at 18	33%	22%	29%	32%
NEET at 18	16%	6%	9%	11%
Russell Group / Oxbridge	11%	27%	21%	23%
Other institution	89%	73%	79%	77%
STEM at uni	39%	44%	35%	50%
Non-STEM at uni	61%	57%	65%	50%
SIC	609	643	680	560
Parental SOC	37	30	35	35
North East	6%	4%	5%	5%
North West	15%	12%	13%	14%
Yorkshire and the Humber	13%	9%	11%	10%
East Midlands	8%	8%	8%	9%
West Midlands	13%	11%	11%	11%
East of England	10%	10%	10%	11%
London	19%	21%	20%	17%
South East	10%	16%	14%	14%
South West	7%	8%	7%	9%
Rest of UK	1%	1%	1%	1%
Job to help people	50%	50%	51%	51%
Job that pays well	54%	51%	52%	53%
Job to be own boss	57%	62%	64%	56%
Job to have non-routine	52%	56%	52%	55%
Job good promotion chances	52%	52%	51%	53%
Job with regular hours	52%	50%	51%	50%
Academic self-concept high	24%	29%	26%	26%
Academic self-concept low	21%	14%	18%	17%
No IAG used in last year	15%	12%	11%	16%
Friends and relatives	55%	66%	64%	58%
Teachers / Lecturers	20%	18%	19%	18%
Connexions	2%	1%	1%	2%
Job centre plus	5%	1%	2%	3%

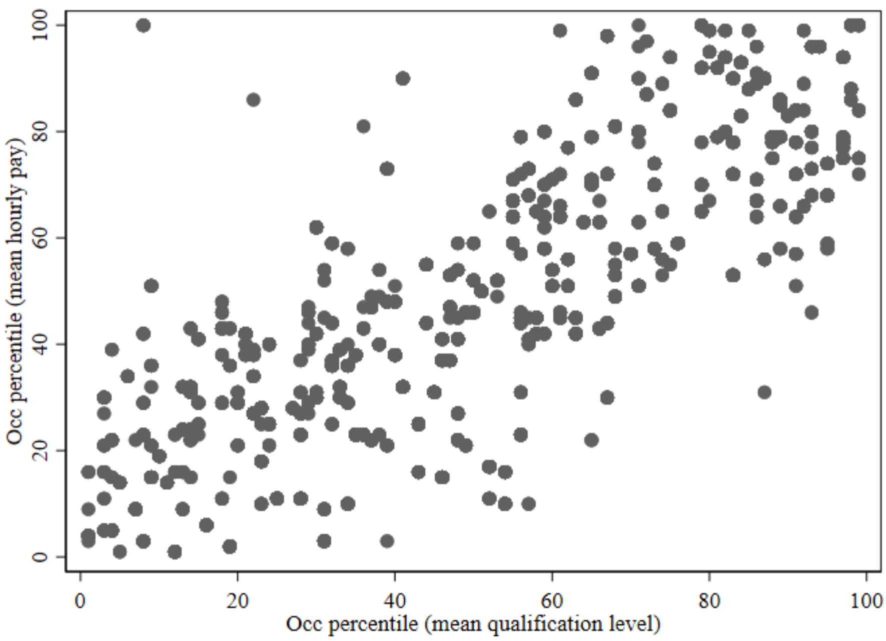
Notes: Source, Next Steps. Samples are N=1285, 1967, 2548, 2196 in columns 1-4. Our sample is defined as survey respondents with both a GCSE total points score from the matched NPD data and a 4-digit occupational SOC code at age 25. Wave 8 final weights applied.

Figure A1 Difference in relative size of occupations between the LFS and Next Steps



Notes: Source, Next Steps. There are 335 4-digit occupation codes in our sample. Our sample is defined as survey respondents with both a GCSE total points score from the matched NPD data and a 4-digit occupational SOC code at age 25.

Figure A2 Relationship between earnings-based occupation ranking and education-based occupation ranking



Notes: Source, Next Steps. Each dot represents an occupations category. There are 335 4-digit occupation codes in our sample. Our sample is defined as survey respondents with both a GCSE total points score from the matched NPD data and a 4-digit occupational SOC code at age 25.

Table A1: Mismatch in education-based and earnings-based match across quintiles of achievement by SES (parental NS-SEC) and gender

	Education match				
Achievement quintiles	1	2	3	4	5
Medium SES	-3.34 (2.71)	-4.29 (1.93)**	-0.75 (1.69)	-0.15 (1.83)	-5.98 (1.92)***
Low SES	-3.67 (2.52)	-8.38 (2.01)***	-5.21 (1.88)***	-3.55 (2.15)	-10.84 (2.22)***
Women	2.03 (1.91)	5.19 (1.59)***	6.47 (1.45)***	3.24 (1.60)**	3.06 (1.57)*
Controls	x	x	x	x	x
N	518	896	1,215	1,086	1,029
	Earning match				
Achievement quintiles	1	2	3	4	5
Medium SES	-2.61 (2.75)	-3.13 (1.91)	-0.20 (1.70)	0.83 (1.82)	-5.04 (1.89)***
Low SES	-2.08 (2.56)	-6.48 (1.99)***	-5.17 (1.90)***	-1.69 (2.14)	-11.12 (2.20)***
Women	-15.31 (1.95)***	-9.62 (1.58)***	-4.80 (1.46)***	-4.78 (1.59)***	-3.54 (1.54)**
Controls	x	x	x	x	x
N	518	896	1,215	1,086	1,029

Notes: Source, Next Steps. Samples reported in final row. Our sample is defined as survey respondents with both a GCSE total points score from the matched NPD data and a 4-digit occupational SOC code at age 25. Wave 8 final weights applied. Low attainers are those in the bottom 20 percentiles of GCSE attainment at age 16. High attainers are those in the top 20 percentiles of GCSE attainment at age 16. Controls include Key Stage 2 points score (age 11) and a categorical measure of ethnicity. Standard errors in parentheses. *** 1% sig, ** 5% sig, * 10% sig.

Table A2: Mismatch in education-based and earnings-based match for those who are low and high attaining by alternative SES measure using parental education

	Education match	
Achievement quintiles	1 - low	5 – high
Medium Ed	0.92	-6.27
	(5.69)	(1.83)***
Low Ed	0.77	-12.62
	(5.67)	(2.30)***
Women	1.52	3.87
	(2.00)	(1.58)***
Controls	x	x
N	488	976
	Earning match	
Achievement quintiles	1 - low	5 – high
Medium Ed	5.02	-6.94
	(5.78)	(1.79)***
Low Ed	6.19	-11.84
	(5.76)	(2.26)***
Women	-16.61	-2.22
	(2.03)***	(1.55)
Controls	x	x
N	488	976

Notes: Source, Next Steps. Samples reported in final row. Our sample is defined as survey respondents with both a GCSE total points score from the matched NPD data and a 4-digit occupational SOC code at age 25. Wave 8 final weights applied. Low attainers are those in the bottom 20 percentiles of GCSE attainment at age 16. High attainers are those in the top 20 percentiles of GCSE attainment at age 16. Controls include Key Stage 2 points score (age 11) and a categorical measure of ethnicity. Standard errors in parentheses. *** 1% sig, ** 5% sig, * 10% sig.

Table A3: Mismatch in earnings-based match, conditioning on later education and participation measures for those who are low and high attaining by SES (parental NS-SEC) and gender

Achievement quintile 1	Baseline	Age 16 to age 18 activity	Age 18 to age 21 activity	All activity 16-21	3 digit SIC	Main parent 2-dig SOC	Destination region	Motivation for occupation choice	Academic self-concept in school	Information, advice and guidance
Medium SES	-2.61 (2.75)	-2.58 (2.75)	-2.64 (2.74)	-2.64 (2.76)	-1.26 (2.45)	-0.30 (2.99)	-3.03 (2.79)	-3.15 (2.77)	-2.65 (2.76)	-2.96 (2.77)
Low SES	-2.08 (2.56)	-1.57 (2.56)	-1.59 (2.55)	-1.27 (2.56)	0.55 (2.25)	-1.71 (2.92)	-2.78 (2.59)	-2.29 (2.56)	-1.57 (2.58)	-2.19 (2.60)
Women	-15.31 (1.95)***	-15.11 (1.94)***	-15.83 (1.94)***	-15.67 (1.95)***	-3.90 (1.97)**	-14.55 (2.00)***	-14.90 (1.95)***	-13.25 (2.03)***	-15.59 (1.96)***	-15.42 (1.96)***
Controls	x	x	x	x	x	x	x	x	x	x
N	518	518	518	518	518	518	518	518	518	518

Achievement quintile 5	Baseline	Age 16 to age 18 activity	Age 18 to age 21 activity	All activity 16-21	3 digit SIC	Main parent 2-dig SOC	Destination region	Motivation for occupation choice at 20	Academic self-concept in school	Information, advice and guidance
Medium SES	-5.04 (1.89)***	-3.97 (1.85)**	-3.36 (1.85)*	-3.07 (1.82)*	-5.17 (1.79)***	-4.70 (2.03)**	-3.90 (1.92)***	-5.41 (1.91)***	-5.22 (1.89)***	-5.06 (1.88)***
Low SES	-11.12 (2.20)***	-8.27 (2.16)***	-8.25 (2.17)***	-7.08 (2.13)***	-8.69 (2.11)***	-9.50 (2.38)***	-8.85 (2.26)***	-11.54 (2.21)***	-11.05 (2.20)***	-11.25 (2.20)***
Women	-3.54 (1.54)**	-3.78 (1.50)**	-3.47 (1.58)**	-3.27 (1.55)**	-3.04 (1.56)*	-3.47 (1.57)**	-3.40 (1.55)**	-3.29 (1.57)**	-3.22 (1.55)**	-4.23 (1.54)***
Controls	x	x	x	x	x	x	x	x	x	x
N	1,029	1,029	1,029	1,029	1,029	1,029	1,029	1,029	1,029	1,029

Notes: Source, Next Steps. Samples reported in final row. Our sample is defined as survey respondents with both a GCSE total points score from the matched NPD data and a 4-digit occupational SOC code at age 25. Wave 8 final weights applied. Low attainers are those in the bottom 20 percentiles of GCSE attainment at age 16. High attainers are those in the top 20 percentiles of GCSE attainment at age 16. Controls include Key Stage 2 points score (age 11) and a categorical measure of ethnicity. Age 16 to age 18 activity: KS5 points scores, not in full time ed age 16-18. Age 18 to 21 activity: In university, working, not in education, employment or training, university institution attended (Oxbridge, RG, other), subject studied (18 JACS principal subject codes). 3 digit SIC: 3-digit industry (SIC code) of the respondent at age 25. Par 2-dig SOC: 2-digit occupation (SOC code) of the main parent in wave 4 (age 16). Destination region: Government office region of the respondent at age 25. Motivation for occupation choice at 20: Statements about what desire from job, including to help other people, a job that pays well, to be my own boss, to have an interesting / non-routine job, to have a chance of promotion, to have regular hours. Academic self-concept: quartiles of academic self-concept score. Information, advice, and guidance: Most useful source of IAG in the past 12 months at age 20, including no IAG used, friends and relatives, teachers and lecturers, connexions, Direct.gov, National Apprenticeship Service, Jobcentre plus advisor, Professional in field of interest, Careers advisors and student support services. Standard errors in parentheses. *** 1% sig, ** 5% sig, * 10% sig.

ucl.ac.uk/ioe/cepeo